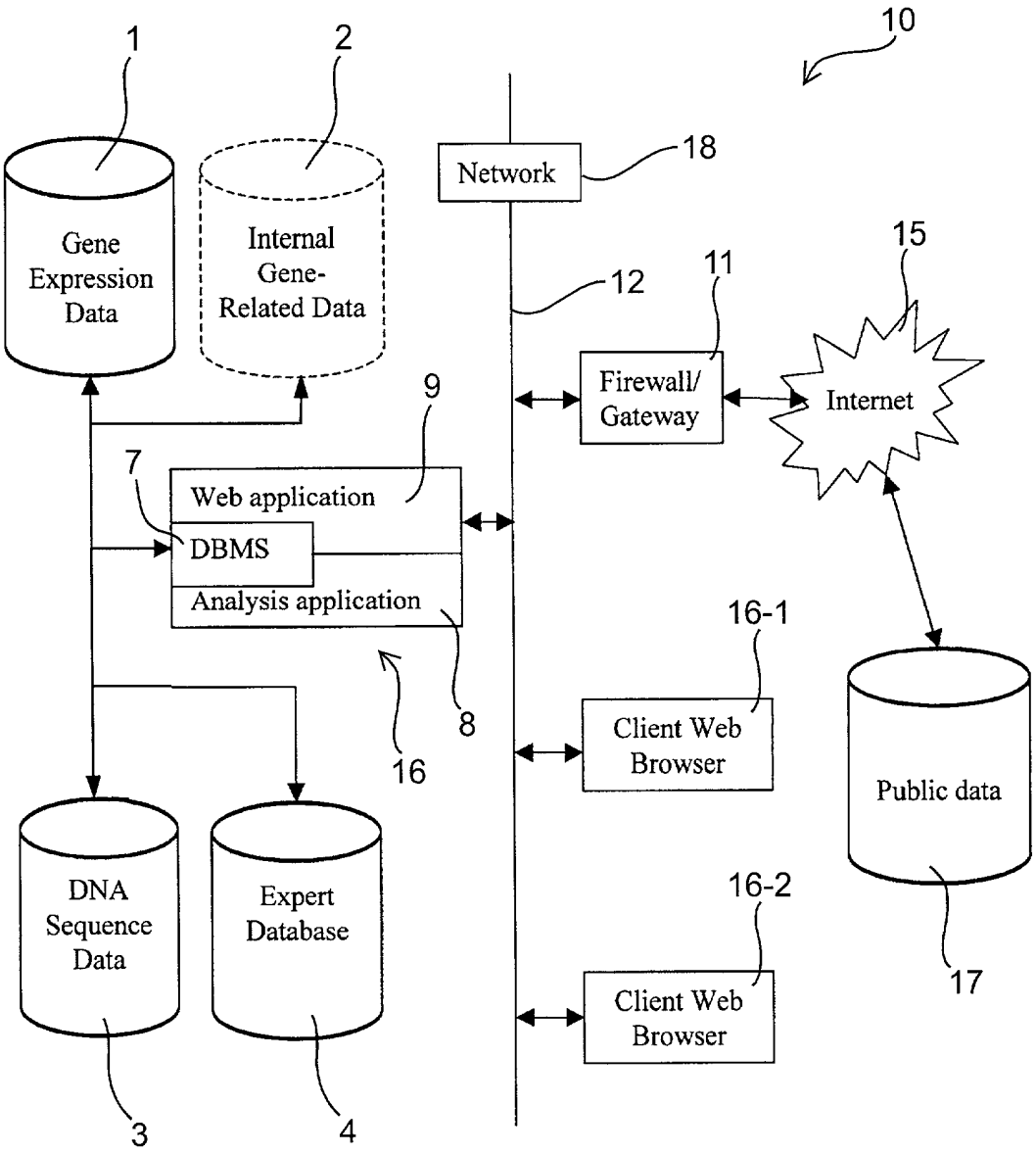
10
1
2
Gene Expression Data
Internal Gene-Related Data
Network
18
12
11
15
Firewall/ Gateway
Internet
9
7
Web application
DBMS
Analysis application
8
16
16-1
Client Web Browser
Public data
17
DNA Sequence Data
Expert Database
16-2
Client Web Browser
3
4

FIG. 1

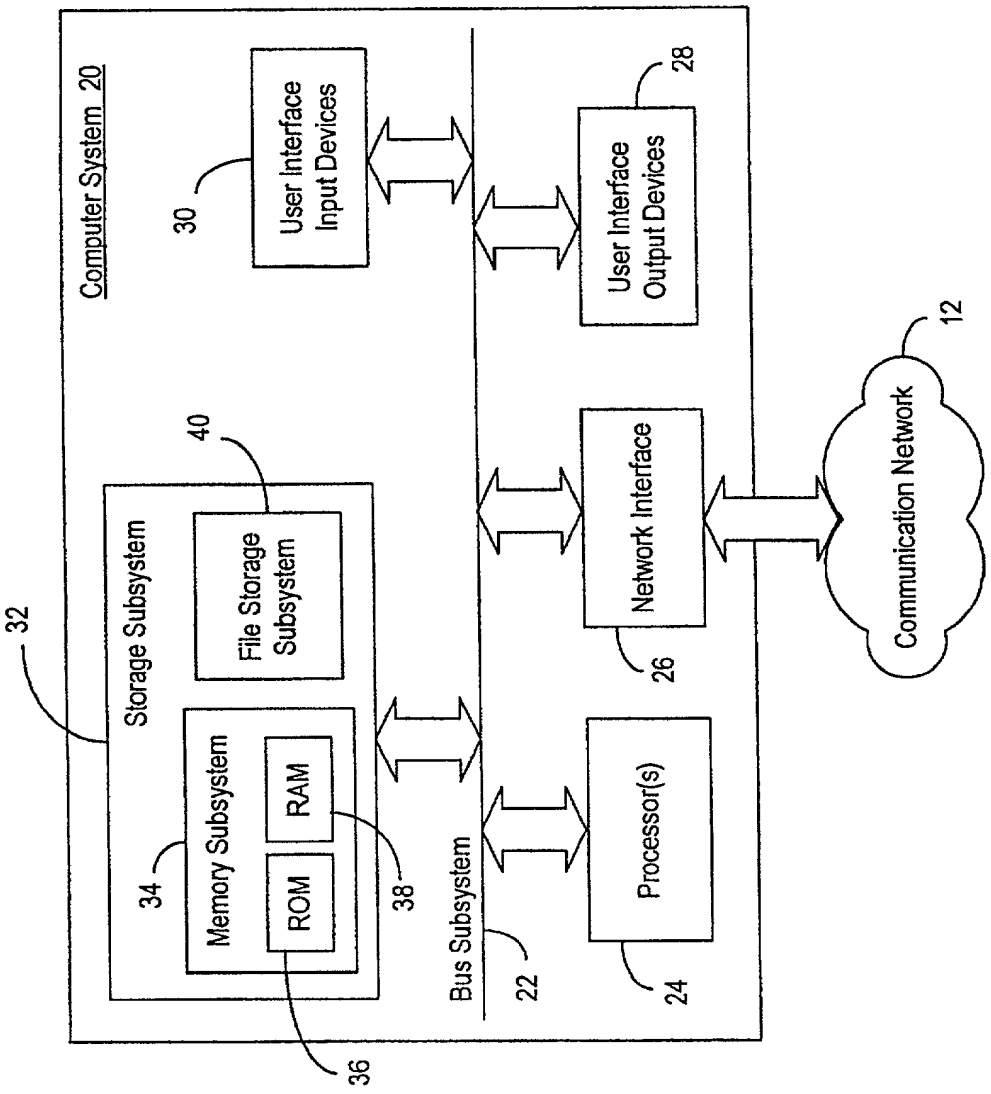
Computer System 20
User Interface Input Devices
30
User Interface Output Devices
28
Storage Subsystem
32
File Storage Subsystem
40
Memory Subsystem
34
RAM
ROM
38
36
Bus Subsystem
22
Processor(s)
24
Network Interface
26
Communication Network
12

FIG. 2

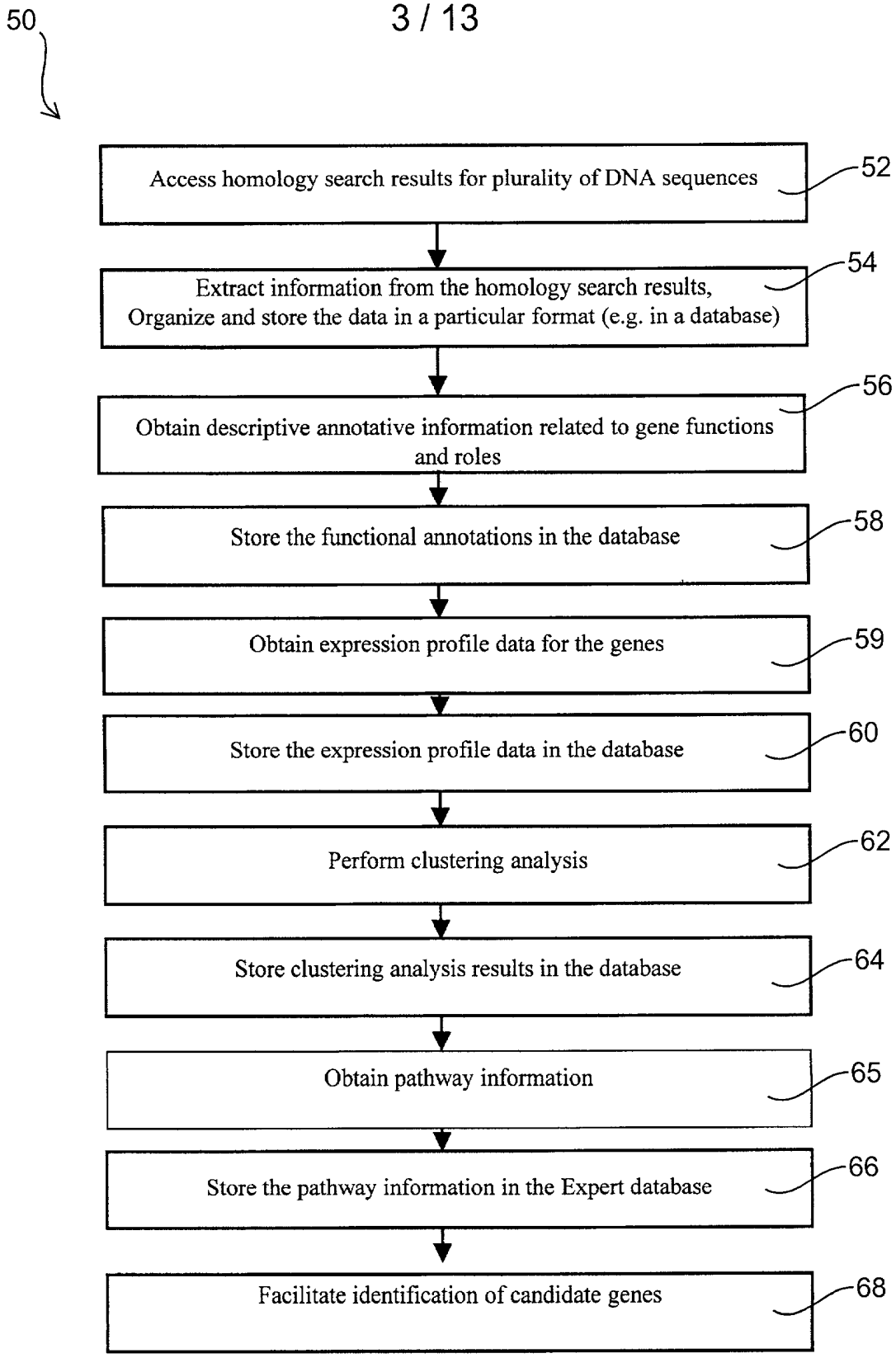
50
Access homology search results for plurality of DNA sequences
52
Extract information from the homology search results, Organize and store the data in a particular format (e.g. in a database)
54
Obtain descriptive annotative information related to gene functions and roles
56
Store the functional annotations in the database
58
Obtain expression profile data for the genes
59
Store the expression profile data in the database
60
Perform clustering analysis
62
Store clustering analysis results in the database
64
Obtain pathway information
65
Store the pathway information in the Expert database
66
Facilitate identification of candidate genes
68

FIG. 3

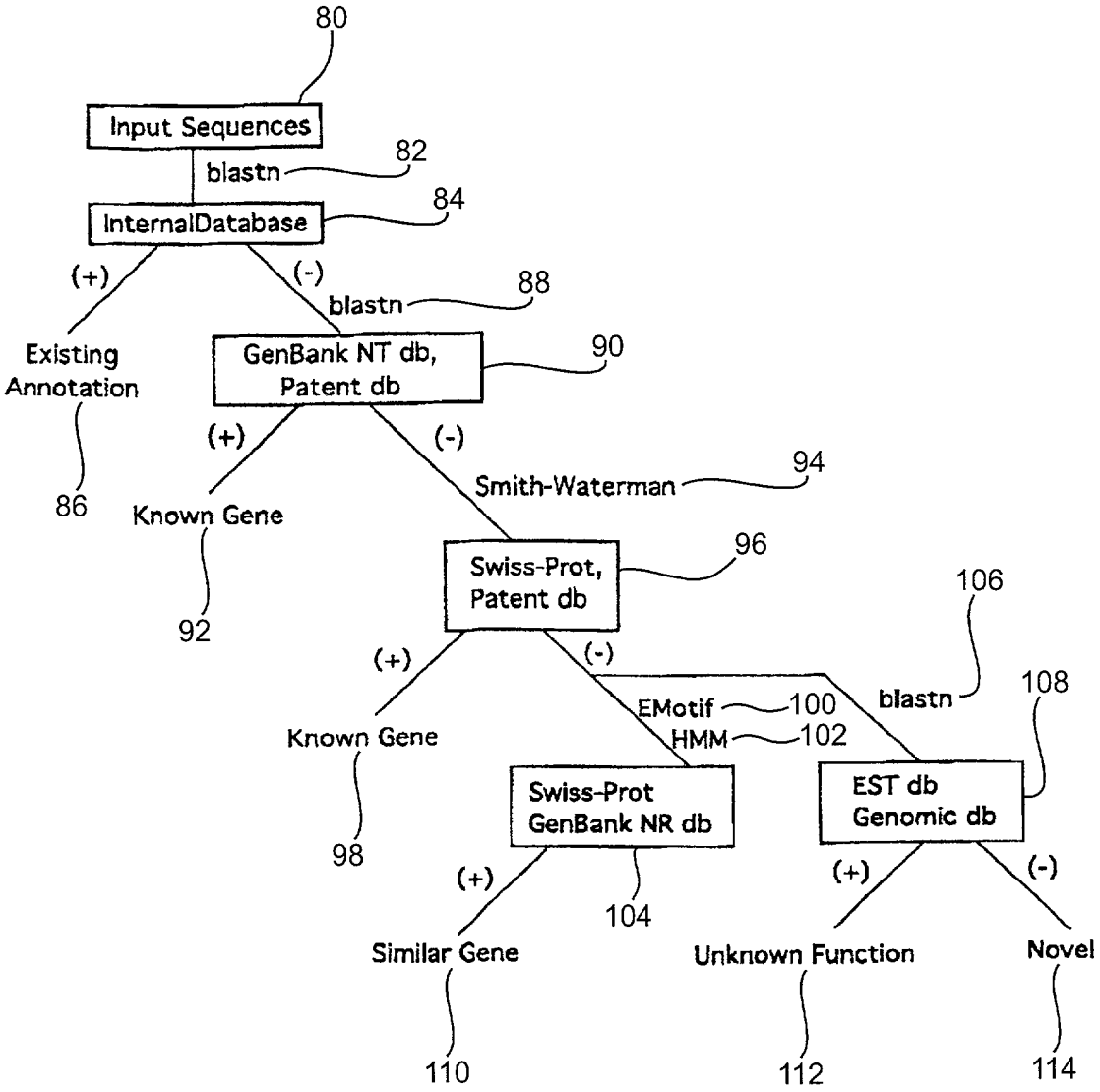
80
Input Sequences
blastn
82
InternalDatabase
84
(+)
(-)
blastn
88
Existing Annotation
GenBank NT db, Patent db
90
86
(+)
(-)
Known Gene
Smith-Waterman
94
92
Swiss-Prot, Patent db
96
106
(+)
(-)
EMotif
100
HMM
102
blastn
108
Known Gene
Swiss-Prot GenBank NR db
EST db Genomic db
98
(+)
(-)
(+)
(-)
104
Similar Gene
Unknown Function
Novel
110
112
114

FIG. 4

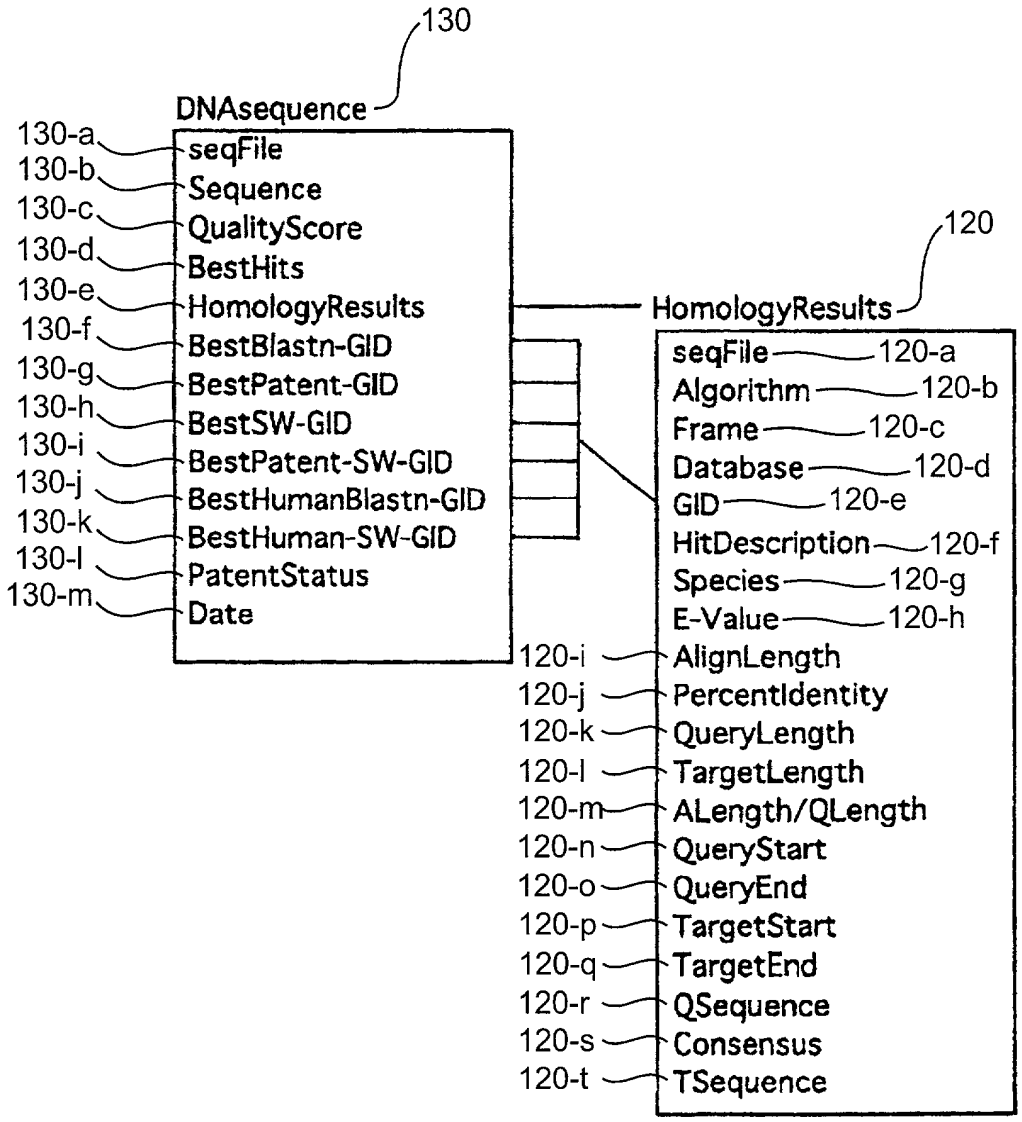
130
DNAsequence
130-a
seqFile
130-b
Sequence
130-c
QualityScore
130-d
BestHits
130-e
HomologyResults
130-f
BestBlastn-GID
130-g
BestPatent-GID
130-h
BestSW-GID
130-i
BestPatent-SW-GID
130-j
BestHumanBlastn-GID
130-k
BestHuman-SW-GID
130-l
PatentStatus
130-m
Date
120
HomologyResults
seqFile 120-a
Algorithm 120-b
Frame 120-c
Database 120-d
GID 120-e
HitDescription 120-f
Species 120-g
E-Value 120-h
120-i
AlignLength
120-j
PercentIdentity
120-k
QueryLength
120-l
TargetLength
120-m
ALength/QLength
120-n
QueryStart
120-o
QueryEnd
120-p
TargetStart
120-q
TargetEnd
120-r
QSequence
120-s
Consensus
120-t
TSequence

FIG. 5

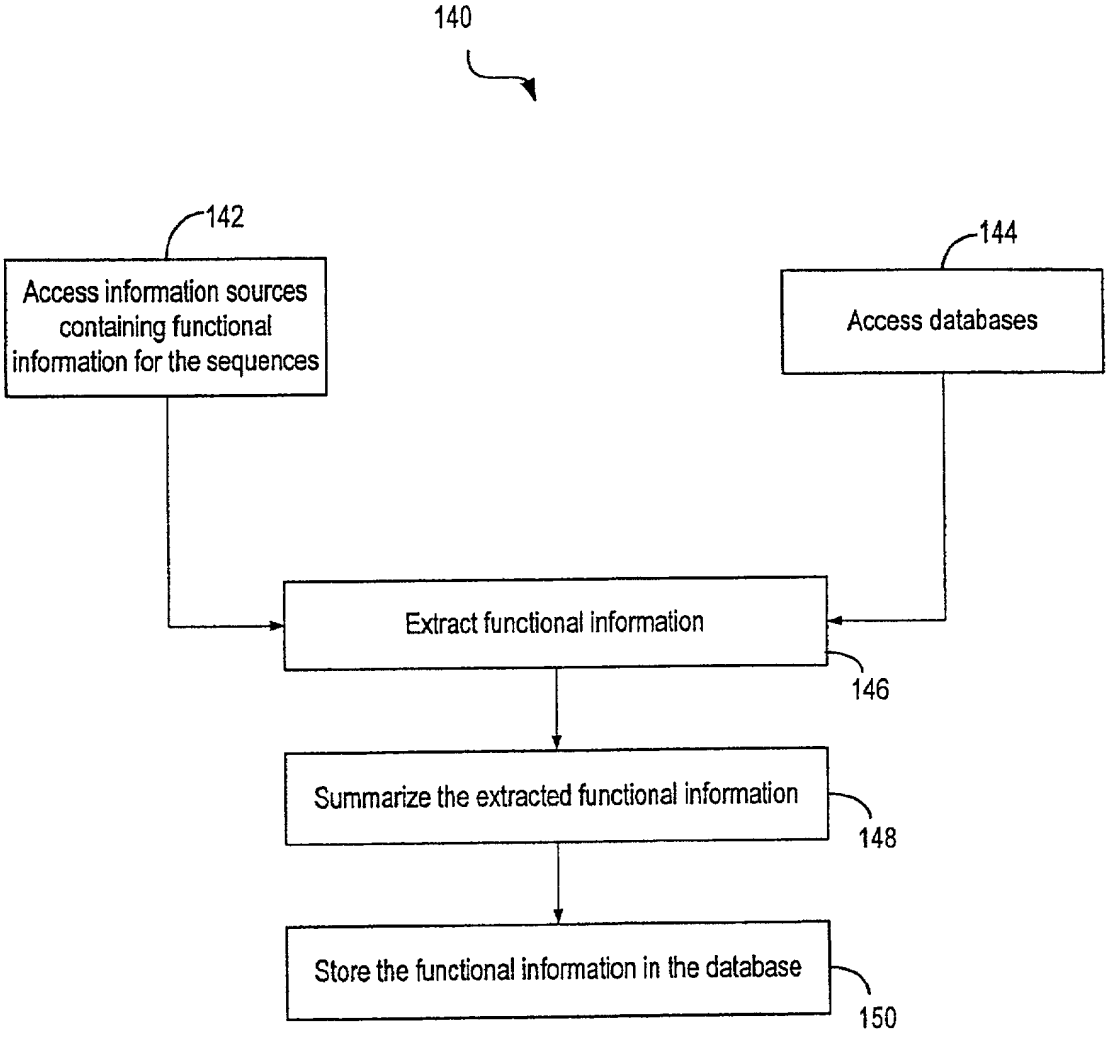
140
142
Access information sources containing functional information for the sequences
144
Access databases
Extract functional information
146
Summarize the extracted functional information
148
Store the functional information in the database
150

FIG. 6

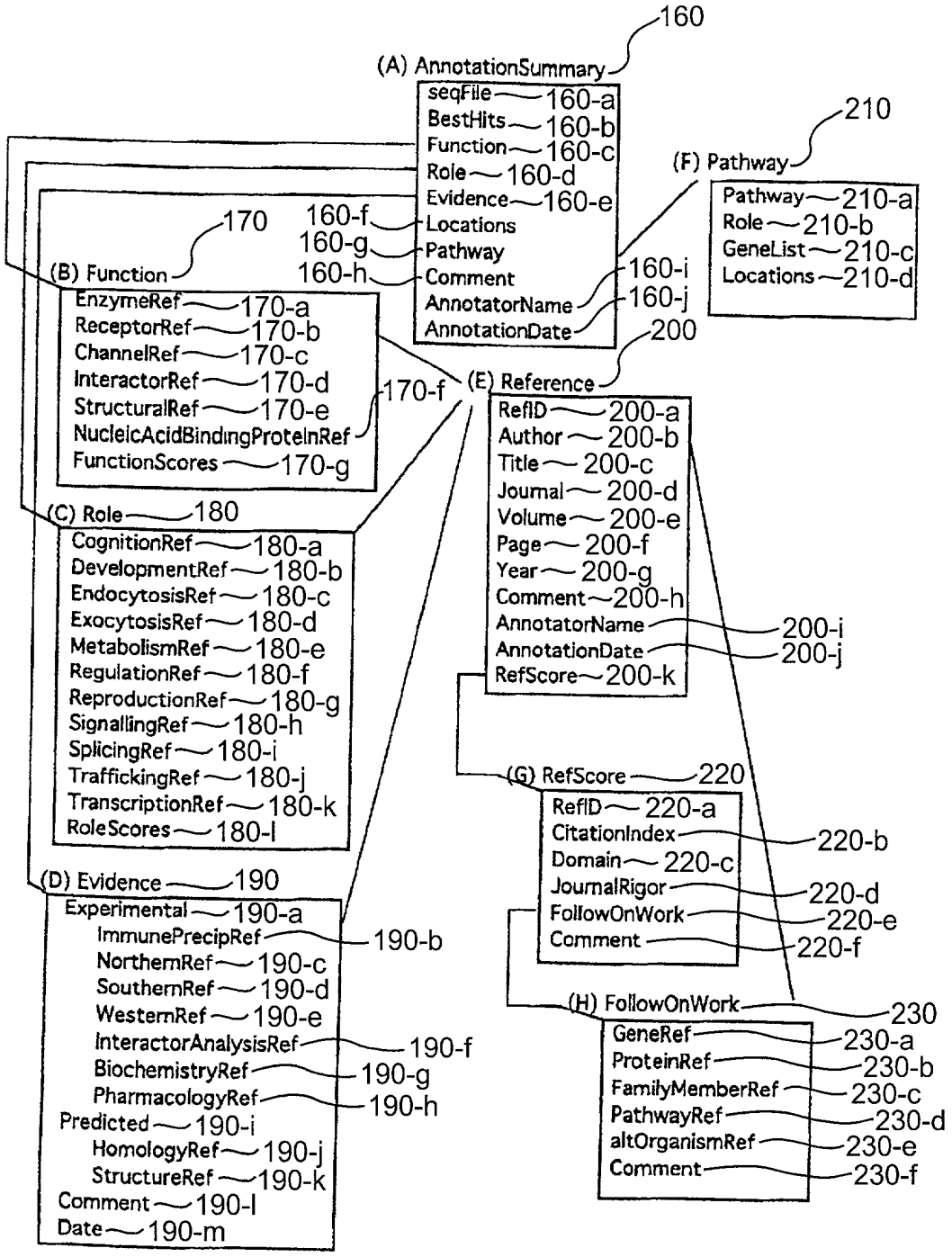
160
(A) AnnotationSummary
seqFile 160-a
BestHits 160-b
Function 160-c
Role 160-d
Evidence 160-e
Locations 160-f
Pathway 160-g
Comment 160-h
AnnotatorName 160-i
AnnotationDate 160-j
210
(F) Pathway
Pathway 210-a
Role 210-b
GeneList 210-c
Locations 210-d
170
(B) Function
EnzymeRef 170-a
ReceptorRef 170-b
ChannelRef 170-c
InteractorRef 170-d
StructuralRef 170-e
NucleicAcidBindingProteinRef 170-f
FunctionScores 170-g
200
(E) Reference
RefID 200-a
Author 200-b
Title 200-c
Journal 200-d
Volume 200-e
Page 200-f
Year 200-g
Comment 200-h
AnnotatorName 200-i
AnnotationDate 200-j
RefScore 200-k
(C) Role 180
CognitionRef 180-a
DevelopmentRef 180-b
EndocytosisRef 180-c
ExocytosisRef 180-d
MetabolismRef 180-e
RegulationRef 180-f
ReproductionRef 180-g
SignallingRef 180-h
SplicingRef 180-i
TraffickingRef 180-j
TranscriptionRef 180-k
RoleScores 180-l
(G) RefScore 220
RefID 220-a
CitationIndex 220-b
Domain 220-c
JournalRigor 220-d
FollowOnWork 220-e
Comment 220-f
(D) Evidence 190
Experimental 190-a
ImmunePrecipRef 190-b
NorthernRef 190-c
SouthernRef 190-d
WesternRef 190-e
InteractorAnalysisRef 190-f
BiochemistryRef 190-g
PharmacologyRef 190-h
Predicted 190-i
HomologyRef 190-j
StructureRef 190-k
Comment 190-l
Date 190-m
(H) FollowOnWork 230
GeneRef 230-a
ProteinRef 230-b
FamilyMemberRef 230-c
PathwayRef 230-d
altOrganismRef 230-e
Comment 230-f

FIG. 7

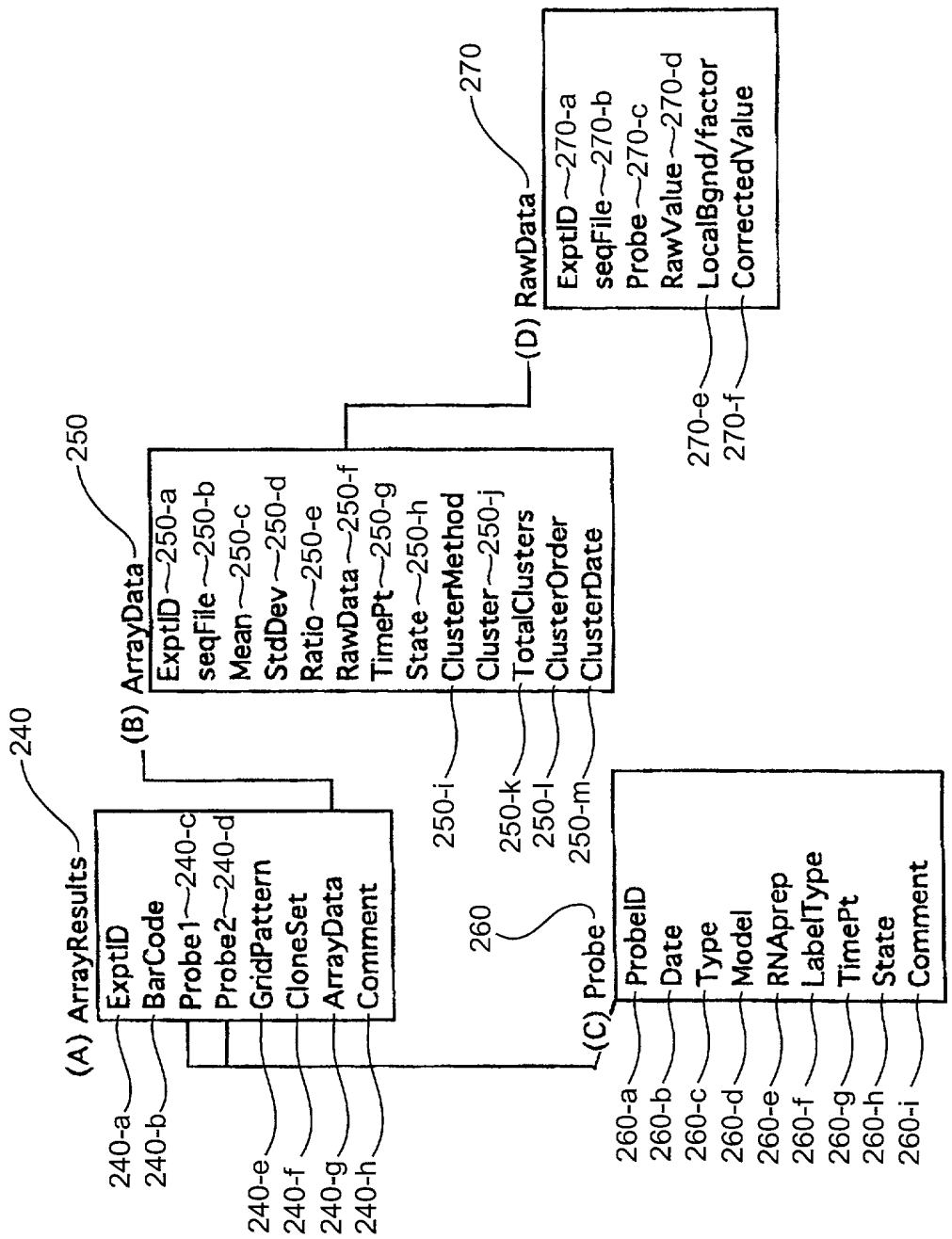
(A) ArrayResults
240
ExptID
BarCode
Probe1
Probe2
GridPattern
CloneSet
ArrayData
Comment
240-a
240-b
240-e
240-f
240-g
240-h
(B) ArrayData
250
ExptID
seqFile
Mean
StdDev
Ratio
RawData
TimePt
State
ClusterMethod
Cluster
TotalClusters
ClusterOrder
ClusterDate
250-i
250-k
250-l
250-m
(C) Probe
260
ProbeID
Date
Type
Model
RNAprep
LabelType
TimePt
State
Comment
260-a
260-b
260-c
260-d
260-e
260-f
260-g
260-h
260-i
(D) RawData
270
ExptID
seqFile
Probe
RawValue
LocalBgnd/factor
CorrectedValue
270-e
270-f

FIG. 8

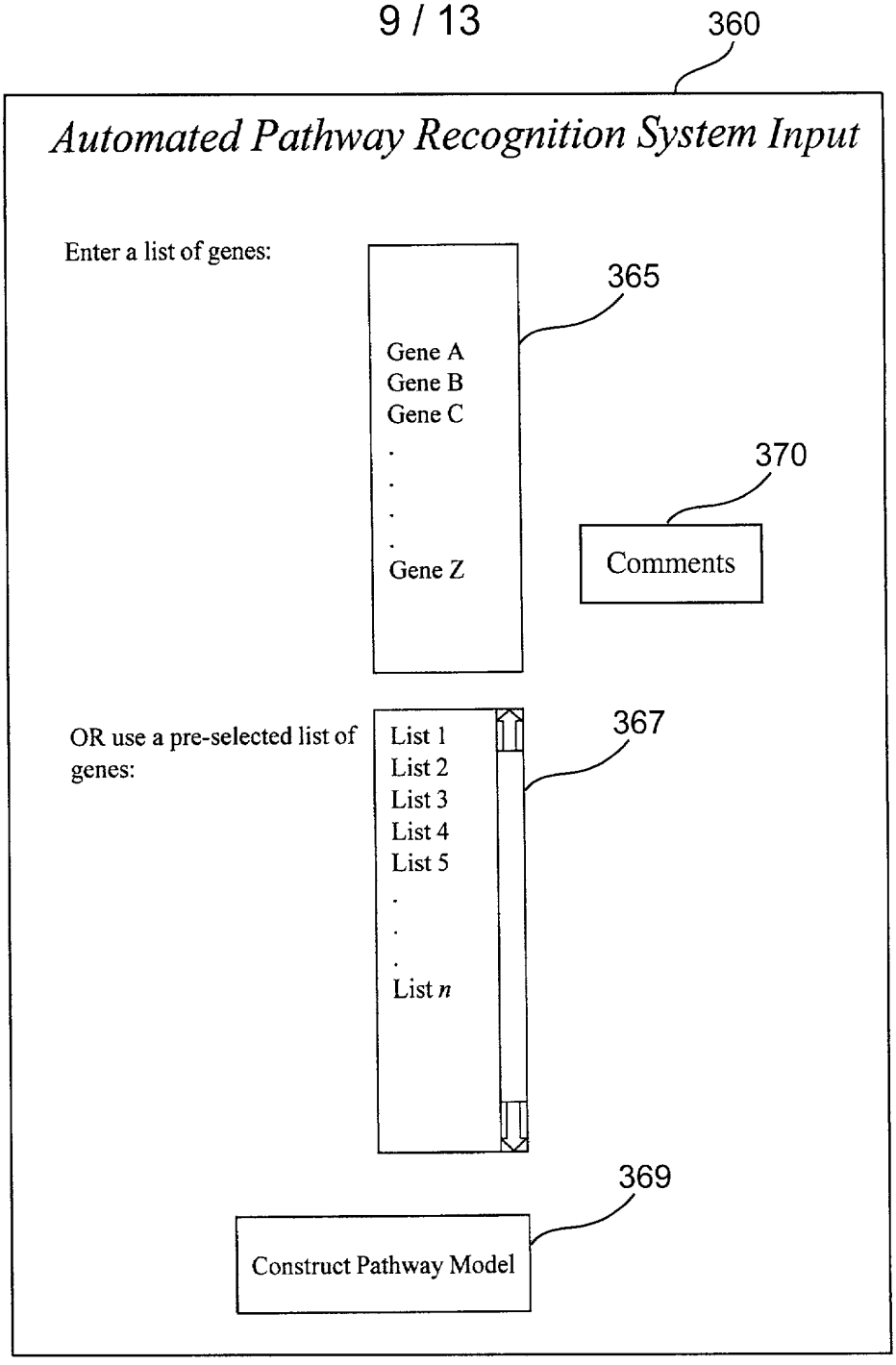
360
Automated Pathway Recognition System Input
Enter a list of genes:
365
Gene A
Gene B
Gene C
Gene Z
370
Comments
OR use a pre-selected list of genes:
367
List 1
List 2
List 3
List 4
List 5
List n
369
Construct Pathway Model

FIG. 9

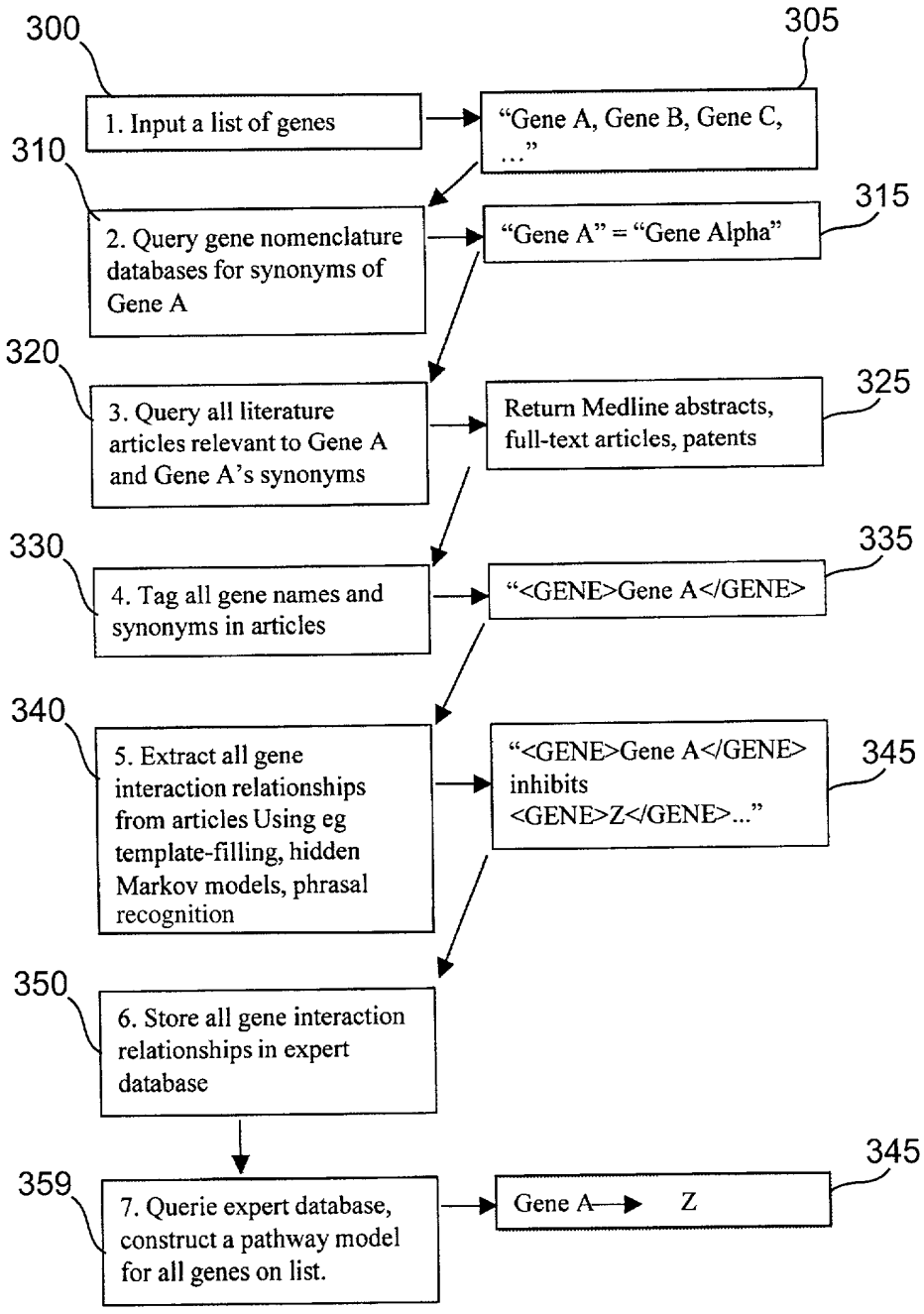
300
1. Input a list of genes
305
"Gene A, Gene B, Gene C, ..."
310
2. Query gene nomenclature databases for synonyms of Gene A
315
"Gene A" = "Gene Alpha"
320
3. Query all literature articles relevant to Gene A and Gene A's synonyms
325
Return Medline abstracts, full-text articles, patents
330
4. Tag all gene names and synonyms in articles
335
"<GENE>Gene A</GENE>
340
5. Extract all gene interaction relationships from articles Using eg template-filling, hidden Markov models, phrasal recognition
345
"<GENE>Gene A</GENE> inhibits <GENE>Z</GENE>..."
350
6. Store all gene interaction relationships in expert database
359
7. Querie expert database, construct a pathway model for all genes on list.
345
Gene A → Z

FIG. 10

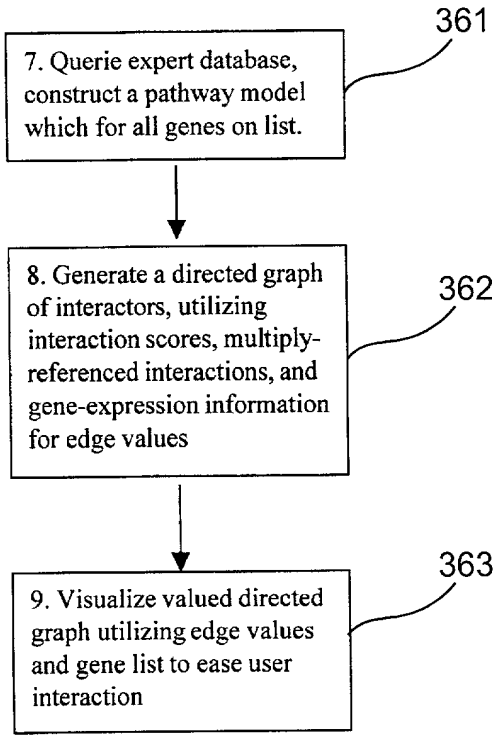
7. Querie expert database, construct a pathway model which for all genes on list.
361
8. Generate a directed graph of interactors, utilizing interaction scores, multiply-referenced interactions, and gene-expression information for edge values
362
9. Visualize valued directed graph utilizing edge values and gene list to ease user interaction
363

FIG. 11

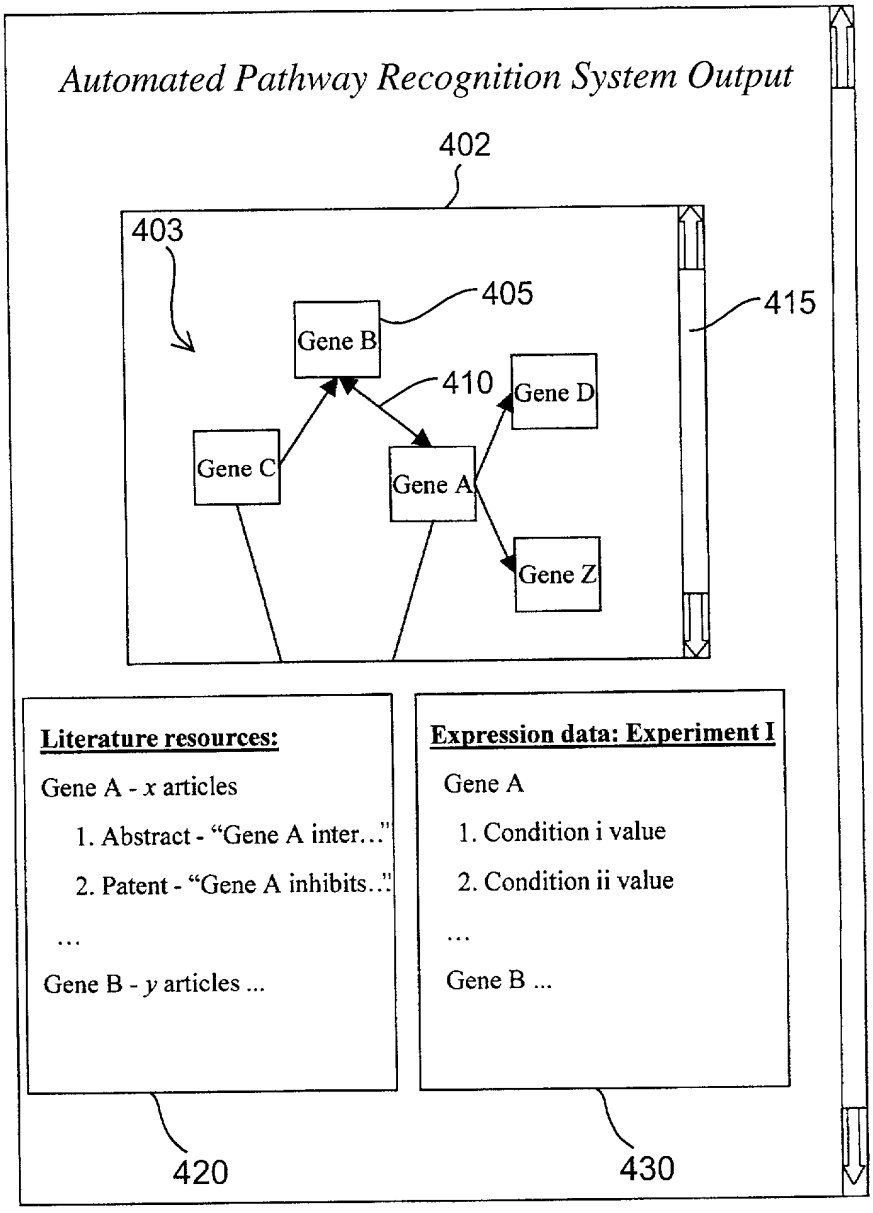
Automated Pathway Recognition System Output
402
403
405
415
Gene B
410
Gene D
Gene C
Gene A
Gene Z
Literature resources:
Gene A - x articles
1. Abstract - "Gene A inter..."
2. Patent - "Gene A inhibits..."
...
Gene B - y articles ...
Expression data: Experiment I
Gene A
1. Condition i value
2. Condition ii value
...
Gene B ...
420
430

FIG. 12

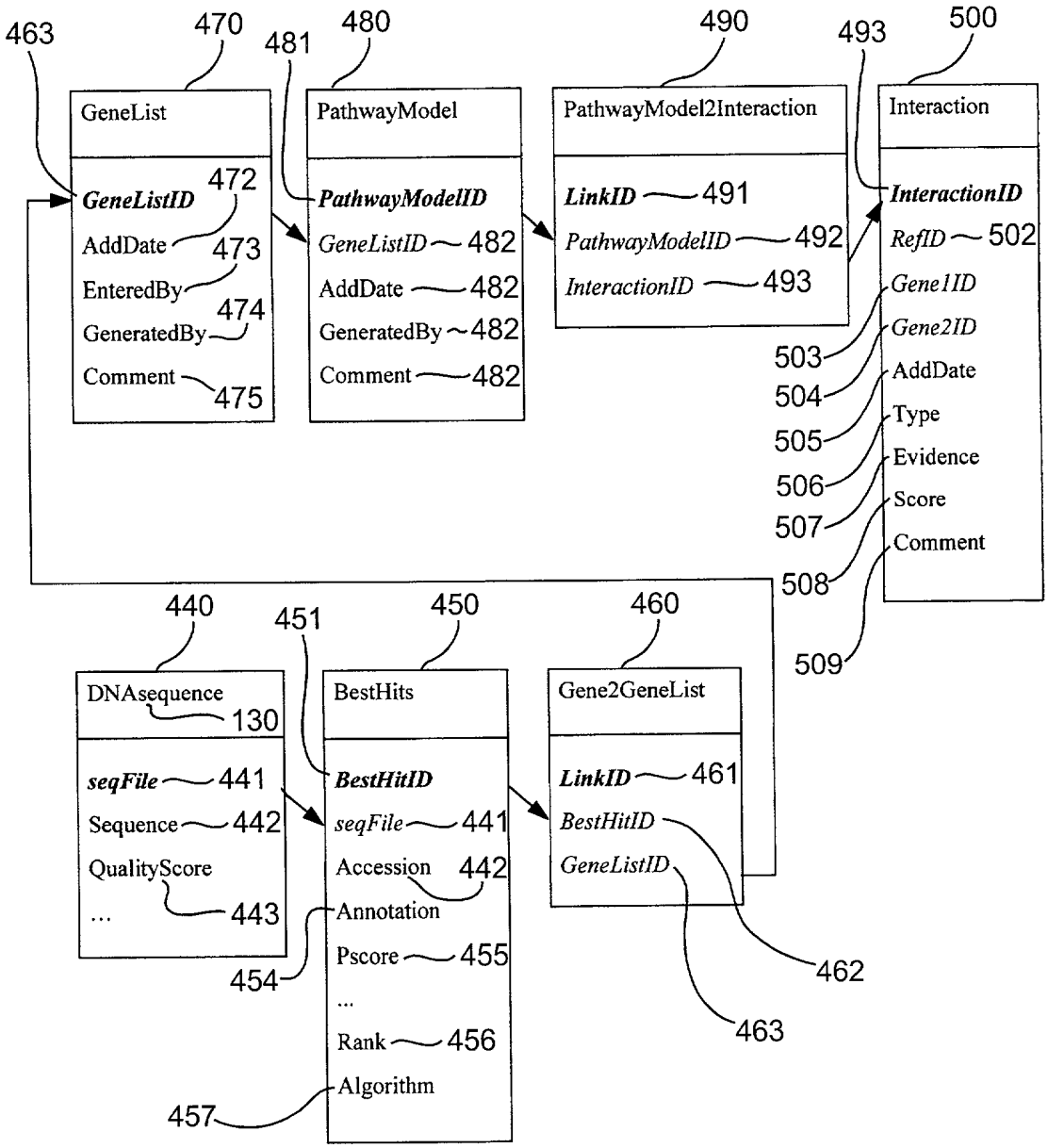
463
470
481
480
490
493
500
GeneList
GeneListID
472
AddDate
473
EnteredBy
474
GeneratedBy
Comment
475
PathwayModel
PathwayModelID
GeneListID ~ 482
AddDate ~ 482
GeneratedBy ~ 482
Comment ~ 482
PathwayModel2Interaction
LinkID ~ 491
PathwayModelID ~ 492
InteractionID ~ 493
Interaction
InteractionID
RefID ~ 502
Gene1ID
Gene2ID
AddDate
Type
Evidence
Score
Comment
503
504
505
506
507
508
509
440
451
450
460
DNAsequence
130
seqFile ~ 441
Sequence ~ 442
QualityScore
...
443
BestHits
BestHitID
seqFile ~ 441
Accession
442
Annotation
Pscore ~ 455
...
Rank ~ 456
Algorithm
454
457
Gene2GeneList
LinkID ~ 461
BestHitID
GeneListID
462
463

FIG. 13